

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

27.12.02

XX

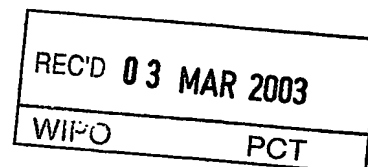
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年12月27日



出 願 番 号

Application Number:

特願2001-396194

[ST.10/C]:

[JP2001-396194]

出 願 人

Applicant(s):

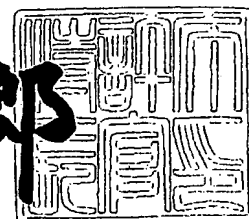
株式会社 先端科学技術インキュベーションセンター

**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

2003年 2月12日

特許庁長官  
Commissioner,  
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3005906

【書類名】 特許願  
【整理番号】 PK010084  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 17/00  
G06F 13/00

【発明者】

【住所又は居所】 神奈川県横浜市都筑区南山田 1-3-4-1202

【氏名】 森下 真一

【発明者】

【住所又は居所】 東京都文京区本郷 5-14-8-302

【氏名】 山田 智之

【特許出願人】

【識別番号】 500524969

【氏名又は名称】 森下 真一

【特許出願人】

【住所又は居所】 東京都文京区本郷 5-14-8-302

【氏名又は名称】 山田 智之

【代理人】

【識別番号】 100103632

【弁理士】

【氏名又は名称】 窪田 英一郎

【選任した代理人】

【識別番号】 100099715

【弁理士】

【氏名又は名称】 吉田 聡

【手数料の表示】

【予納台帳番号】 058377

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 配列の出現頻度算出方法、孤立度算出方法、プライマーへの適正度推定方法および可視化方法、並びに、プログラム

【特許請求の範囲】

【請求項 1】 ゲノム配列中の配列の出現頻度を示す指標を算出する方法であって、

所定長の部分列の各々に関して、前記ゲノム配列中の出現頻度を算出するステップと、

前記所定長の部分列に関する出現頻度を、頻度テーブルに記憶するステップとを備えたことを特徴とする方法。

【請求項 2】 前記頻度テーブルに記憶するステップが、

出現頻度が 0（ゼロ）である部分列に関して、頻度テーブルへの記憶を省略するステップと、

前記所定長より短い第 2 の所定長を有するような第 2 の部分列の各々を用いて、前記第 2 の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を、第 2 のテーブルに記憶するステップとを有することを特徴とする請求項 1 に記載の方法。

【請求項 3】 ゲノム配列中の配列の孤立度を示す指標を算出する方法であって、

所定長の部分列の各々に関して、 $j$  個の塩基を変換した  $j$  ミューテーション（ $j = 1, 2, \dots, i - 1$ ）が、前記ゲノム配列中には表れないが、 $i$  個の塩基を変換した  $i$  ミューテーションが、前記ゲノム配列中に表れるような孤立度  $i$  を算出するステップと、

前記所定長の部分列に関して、前記孤立度を、孤立度テーブルに記憶するステップとを備えたことを特徴とする方法。

【請求項 4】 前記孤立度を算出するステップが、

前記所定長の部分列に関して、当該部分列の  $k$  個の塩基を変換した  $k$  ミューテーションについて、所定長の部分列の各々に関してゲノム配列中の出現頻度を記憶した頻度テーブルを参照して、当該  $k$  ミューテーションが存在するか否かを判

断するステップと、

前記 k ミューテーションが存在する場合に、当該 k を孤立度とするステップと

前記 k ミューテーションが存在しない場合に、k をインクリメントして、前記 k ミューテーションの存在の有無を判断するステップを繰り返し実行するステップとを有することを特徴とする請求項 3 に記載の方法。

【請求項 5】 前記孤立度を算出するステップが、

前記所定長より短い第 2 の所定長を有するような第 2 の部分列の各々を用いて、前記第 2 の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を記憶した第 2 のテーブルを参照して、前記所定長の部分列のうち先頭から第 2 の所定長にある k 個の塩基を変換したような、前記所定長の k ミューテーションが存在するか否かを判断するステップと、

前記 k ミューテーションが存在する場合に、当該 k ミューテーションと、前記所定長の配列との間のハミング距離を見出すステップと、

前記ハミング距離の最小値が k である場合に、当該 k を孤立度とするステップと、

最小値が k より大きい場合には、k をインクリメントして、前記所定長の k ミューテーションの存在の有無、および、ハミング距離の最小値を用いた判断をなすステップを繰り返し実行するステップを有することを特徴とする請求項 3 または 4 に記載の方法。

【請求項 6】 請求項 1 または 2 に記載の方法を用いて生成された頻度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 1 の指標を算出する方法であって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、

前記特定された出現頻度の総和に基づき、第 1 の指標を算出するステップとを備えたことを特徴とする方法。

【請求項 7】 請求項 3 ないし 5 の何れか一項に記載の方法を用いて生成された孤立度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 2 の指標を算出する方法であって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、

前記特定された孤立度の総和に基づき、第 2 の指標を算出するステップとを備えたことを特徴とする方法。

【請求項 8】 請求項 1 または 2 に記載の方法を用いて生成された頻度テーブル、および、請求項 3 ないし 5 の何れか一項に記載の方法を用いて生成された孤立度テーブルを用いて、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 3 の指標を算出する方法であって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、

前記特定された出現頻度の総和に基づき、第 1 の指標を算出するステップと、

前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、

前記特定された孤立度の総和に基づき、第 2 の指標を算出するステップとを備えたことを特徴とする方法。

【請求項 9】 請求項 6 ないし 8 の何れか一項に記載の方法を用いて得られた指標に基づき、当該指標の値或いは範囲にしたがって異なる表示態様を割り当てるステップと、

前記割り当てられた表示態様にしたがって、ゲノム配列中の各塩基を表したような画像を生成するステップとを備えたことを特徴とする方法。

【請求項 10】 前記表示態様が、色彩であることを特徴とする請求項 9 に記載の方法。

【請求項 1 1】 ゲノム配列中の配列の出現頻度を示す指標を算出するためにコンピュータを動作させる、当該コンピュータにより読み取り可能なプログラムであって、

所定長の部分列の各々に関して、前記ゲノム配列中の出現頻度を算出するステップと、

前記所定長の部分列に関する出現頻度を、頻度テーブルに記憶するステップとを前記コンピュータに実行させることを特徴とするプログラム。

【請求項 1 2】 前記頻度テーブルに記憶するステップにおいて、

出現頻度が 0（ゼロ）である部分列に関して、頻度テーブルへの記憶を省略するステップと、

前記所定長より短い第 2 の所定長を有するような第 2 の部分列の各々を用いて、前記第 2 の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を、第 2 のテーブルに記憶するステップとを、前記コンピュータに実行させることを特徴とする請求項 1 1 に記載のプログラム。

【請求項 1 3】 ゲノム配列中の配列の孤立度を示す指標を算出するためにコンピュータを動作させる、当該コンピュータにより読み取り可能なプログラムであって、

所定長の部分列の各々に関して、 $j$  個の塩基を変換した  $j$  ミューテーション（ $j = 1, 2, \dots, i - 1$ ）が、前記ゲノム配列中には表れないが、 $i$  個の塩基を変換した  $i$  ミューテーションが、前記ゲノム配列中に表れるような孤立度  $i$  を算出するステップと、

前記所定長の部分列に関して、前記孤立度を、孤立度テーブルに記憶するステップとを、前記コンピュータに実行させることを特徴とするプログラム。

【請求項 1 4】 前記孤立度を算出するステップにおいて、

前記所定長の部分列に関して、当該部分列の  $k$  個の塩基を変換した  $k$  ミューテーションについて、所定長の部分列の各々に関してゲノム配列中の出現頻度を記憶した頻度テーブルを参照して、当該  $k$  ミューテーションが存在するか否かを判断するステップと、

前記  $k$  ミューテーションが存在する場合に、当該  $k$  を孤立度とするステップと

前記  $k$  ミューテーションが存在しない場合に、 $k$  をインクリメントして、前記  $k$  ミューテーションの存在の有無を判断するステップを繰り返し実行するステップとを、前記コンピュータに実行させることを特徴とする請求項 13 に記載のプログラム。

【請求項 15】 前記孤立度を算出するステップにおいて、

前記所定長より短い第 2 の所定長を有するような第 2 の部分列の各々を用いて、前記第 2 の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を記憶した第 2 のテーブルを参照して、前記所定長の部分列のうち先頭から第 2 の所定長にある  $k$  個の塩基を変換したような、前記所定長の  $k$  ミューテーションが存在するか否かを判断するステップと、

前記  $k$  ミューテーションが存在する場合に、当該  $k$  ミューテーションと、前記所定長の配列との間のハミング距離を見出すステップと、

前記ハミング距離の最小値が  $k$  である場合に、当該  $k$  を孤立度とするステップと、

最小値が  $k$  より大きい場合には、 $k$  をインクリメントして、前記所定長の  $k$  ミューテーションの存在の有無、および、ハミング距離の最小値を用いた判断をなすステップを繰り返し実行するステップを、前記コンピュータに実行させることを特徴とする請求項 13 または 14 に記載のプログラム。

【請求項 16】 請求項 11 または 12 に記載のプログラムを、コンピュータに実行させることにより生成された頻度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 1 の指標を算出するために、コンピュータを動作させる、コンピュータにより読み取り可能なプログラムであって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、

前記特定された出現頻度の総和に基づき、第 1 の指標を算出するステップとを



、前記コンピュータに実行させることを特徴とするプログラム。

【請求項 1 7】 請求項 1 3 ないし 1 5 の何れか一項に記載のプログラムを、コンピュータに実行させることにより生成された孤立度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 2 の指標を算出するために、コンピュータを動作させる、コンピュータにより読み取り可能なプログラムであって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、

前記特定された孤立度の総和に基づき、第 2 の指標を算出するステップとを、前記コンピュータに実行させることを特徴とするプログラム。

【請求項 1 8】 請求項 1 1 または 1 2 に記載のプログラムを、コンピュータに実行させることにより生成された頻度テーブル、および、請求項 1 3 ないし 1 5 の何れか一項に記載のプログラムを、コンピュータに実行させることにより生成された孤立度テーブルを用いて、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第 3 の指標を算出するために、コンピュータを動作させる、コンピュータにより読み取り可能なプログラムであって、

ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、

前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、

前記特定された出現頻度の総和に基づき、第 1 の指標を算出するステップと、

前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、

前記特定された孤立度の総和に基づき、第 2 の指標を算出するステップとを、前記コンピュータにより実行させることを特徴とするプログラム。

【請求項 1 9】 請求項 1 6 ないし 1 8 の何れか一項に記載のプログラムを、コンピュータに実行させることにより、得られた指標に基づき、当該指標の値或

いは範囲にしたがって異なる表示態様を割り当てるステップと、

前記割り当てられた表示態様にしたがって、ゲノム配列中の各塩基を表したような画像を生成するステップとを、前記コンピュータに実行させることを特徴とするプログラム。

#### 【発明の詳細な説明】

#### 【0001】

#### 【発明の属する技術分野】

本発明は、プライマーの選択を支援する手法に関する。

#### 【0002】

#### 【従来の技術】

これまでに多くのプライマーデザインの方法が提案されたが、一つの場所にしかアニールしないようなプライマーを設計するのは困難なのが現状である。データベースに登録されたEST配列に対して、比較的短い配列（K-タプル：K-tuple）のあらゆる可能な塩基配列の組み合わせの出現回数を計算すること、たとえば、 $4^8$ （65536）種類ある8-merの塩基配列の出現回数を計算することで、高頻度に出てくる配列と、低頻度で出てくる配列を見つけることができる。このような手法は、たとえば、「Nucleic Acids Res. 19 3887-3891 (R. Griffais, P. M. AndreおよびM. Thibon: 1991年)」に開示されている。

#### 【0003】

#### 【発明が解決しようとする課題】

しかしながら、いくつかの著名なESTデータベースには、多くの研究者から投稿された類似配列が多数存在するため、このままでは配列の出現頻度を議論することができない。

また、遺伝子をはさむようなプライマーを設計する場合には、プロモーター領域の配列が必要となるため、ESTデータベースだけではプライマーを設計することができないという問題点があった。

#### 【0004】

DNAポリメラーゼは、いくつかのミスマッチが入ったようなオリゴヌクレオチドであっても、プライマーとして認識することが知られている（たとえば、「

Molecular Biology Vol. 28 No. 5, Part I 661-663 (L. B. D&#0;yachenko, A. A. Chenchick, G. L. Khaspekov, A.O. TatarenkoおよびR. Sh. Bibilashvili: 1994年) 参照)。しかしながら、これまでのプライマーデザインの方法においては、ゲノムワイドなミスマッチ耐性について考慮されていなかった。また、これまでのプライマーデザインの方法においては、プライマーの塩基配列を決定した後に、ミスマッチ耐性をデータベースに対して検索していたため、検索に時間がかかるという問題点もあった。

#### 【0005】

本発明は、ユニークなプライマーの設計を支援することを目的とする。

#### 【0006】

##### 【課題を解決するための手段】

本発明においては、ゲノム配列中の所定長 (N-mer) の配列の出現回数を計数するとともに、ミスマッチ耐性を評価する値として、配列特異性のもう一つの側面としての孤立度を導入し、評価した。孤立度は、たとえば、配列同士の最小ハミング距離として定義した。孤立度を導入することにより、塩基配列の特異性をさらに精密に分類できるようになった。

#### 【0007】

より詳細には、本発明の目的は、ゲノム配列中の配列の出現頻度を示す指標を算出する方法であって、所定長の部分列の各々に関して、前記ゲノム配列中の出現頻度を算出するステップと、前記所定長の部分列に関する出現頻度を、頻度テーブルに記憶するステップとを備えたことを特徴とする方法により達成される。上記頻度テーブルに記憶するステップは、出現頻度が0（ゼロ）である部分列に関して、頻度テーブルへの記憶を省略するステップと、前記所定長より短い第2の所定長を有するような第2の部分列の各々を用いて、前記第2の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を、第2のテーブルに記憶するステップとを有するのが、メモリ容量の削減および処理時間の短縮のために望ましい。

#### 【0008】

また、本発明の目的は、ゲノム配列中の配列の孤立度を示す指標を算出する方

法であって、所定長の部分列の各々に関して、 $j$  個の塩基を変換した  $j$  ミューテーション ( $j = 1, 2, \dots, i-1$ ) が、前記ゲノム配列中には表れないが、 $i$  個の塩基を変換した  $i$  ミューテーションが、前記ゲノム配列中に表れるような孤立度  $i$  を算出するステップと、前記所定長の部分列に関して、前記孤立度を、孤立度テーブルに記憶するステップとを備えたことを特徴とする方法によっても達成される。

#### 【0009】

上記出現頻度および／または孤立度を用いて、ゲノム配列において特異性のある部分を特定することが容易になり、よりユニークなプライマーを設計することが可能となる。

#### 【0010】

好ましい実施態様においては、前記孤立度を算出するステップが、前記所定長の部分列に関して、当該部分列の  $k$  個の塩基を変換した  $k$  ミューテーションについて、所定長の部分列の各々に関してゲノム配列中の出現頻度を記憶した頻度テーブルを参照して、当該  $k$  ミューテーションが存在するか否かを判断するステップと、前記  $k$  ミューテーションが存在する場合に、当該  $k$  を孤立度とするステップと、前記  $k$  ミューテーションが存在しない場合に、 $k$  をインクリメントして、前記  $k$  ミューテーションの存在の有無を判断するステップを繰り返し実行するステップとを有する。

#### 【0011】

より好ましい実施態様においては、前記孤立度を算出するステップが、前記所定長より短い第2の所定長を有するような第2の部分列の各々を用いて、前記第2の部分列を先頭から含む所定長の部分列の、前記頻度テーブル中の位置を記憶した第2のテーブルを参照して、前記所定長の部分列のうち先頭から第2の所定長にある  $k$  個の塩基を変換したような、前記所定長の  $k$  ミューテーションが存在するか否かを判断するステップと、前記  $k$  ミューテーションが存在する場合に、当該  $k$  ミューテーションと、前記所定長の配列との間のハミング距離を見出すステップと、前記ハミング距離の最小値が  $k$  である場合に、当該  $k$  を孤立度とするステップと、最小値が  $k$  より大きい場合には、 $k$  をインクリメントして、前記所

定長のkミューテーションの存在の有無、および、ハミング距離の最小値を用いた判断をなすステップを繰り返し実行するステップを有する。

【0012】

また、本発明の目的は、上記方法を用いて生成された頻度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第1の指標を算出する方法であって、ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、前記特定された出現頻度の総和に基づき、第1の指標を算出するステップとを備えたことを特徴とする方法によっても達成される。

【0013】

また、本発明の目的は、上記方法を用いて生成された孤立度テーブルを利用して、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第2の指標を算出する方法であって、ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、前記特定された孤立度の総和に基づき、第2の指標を算出するステップとを備えたことを特徴とする方法によっても達成される。

【0014】

さらに、本発明の目的は、上記方法を用いて生成された頻度テーブル、および、孤立度テーブルを用いて、ゲノム配列中の塩基に関して、当該塩基を含む配列のプライマーへの適性を示す第3の指標を算出する方法であって、ゲノム配列を構成する各塩基に関して、当該塩基を含む所定長と同数の配列を特定するステップと、前記頻度テーブルを参照して、特定された配列の各々に関する出現頻度を特定するステップと、前記特定された出現頻度の総和に基づき、第1の指標を算出するステップと、前記孤立度テーブルを参照して、特定された配列の各々に関する孤立度を特定するステップと、前記特定された孤立度の総和に基づき、第2の指標を算出するステップとを備えたことを特徴とする方法によっても達成される。これらの方法により、ゲノム配列中の塩基レベルでの指標を得ることが可能

となり、よりユニークなプライマーの設計を支援することが可能となる。

#### 【0015】

さらに、本発明の目的は、上記方法を用いて得られた指標に基づき、当該指標の値或いは範囲にしたがって異なる表示態様を割り当てるステップと、前記割り当てられた表示態様にしたがって、ゲノム配列中の各塩基を表したような画像を生成するステップとを備えたことを特徴とする方法によっても達成される。たとえば、上記表示態様は、色彩であっても良い。

#### 【0016】

また、本発明の目的は、ゲノム配列中の配列の出現頻度を示す指標を算出するためにコンピュータを動作させる、当該コンピュータにより読み取り可能なプログラムであって、所定長の部分列の各々に関して、前記ゲノム配列中の出現頻度を算出するステップと、前記所定長の部分列に関する出現頻度を、頻度テーブルに記憶するステップとを前記コンピュータに実行させることを特徴とするプログラムによっても達成される。

#### 【0017】

さらに、本発明の目的は、ゲノム配列中の配列の孤立度を示す指標を算出するためにコンピュータを動作させる、当該コンピュータにより読み取り可能なプログラムであって、所定長の部分列の各々に関して、 $j$  個の塩基を変換した  $j$  ミューテーション ( $j = 1, 2, \dots, i-1$ ) が、前記ゲノム配列中には表れないが、 $i$  個の塩基を変換した  $i$  ミューテーションが、前記ゲノム配列中に表れるような孤立度  $i$  を算出するステップと、前記所定長の部分列に関して、前記孤立度を、孤立度テーブルに記憶するステップとを、前記コンピュータに実行させることを特徴とするプログラムによっても達成される。

#### 【0018】

##### 【発明の実施の形態】

以下、添付図面を参照して、本発明の実施の形態につき説明を加える。図1は、本発明の実施の形態にかかるプライマー設計支援システムの概略を示すブロックダイアグラムである。図1に示すように、このプライマー設計支援システム10は、プライマー設計のために利用するゲノム配列を参照して、当該ゲノム配列

中の、所定長の部分列（N-mer 配列）の出現頻度を算出する出現頻度算出部 12 と、後述するような、ゲノム配列に対する、各部分列の孤立度を算出する孤立度算出部 14 と、各部分列に関する出現頻度および孤立度を記憶する頻度／孤立度テーブル 16 と、頻度／孤立度テーブル 16 を参照して、ゲノム配列を可視化して表示するために必要な処理を実行する可視化処理部 18 と、可視化処理部 18 による処理により表示装置 24 の画面上に表示された画像を参照して、ユーザが入力装置（図示せず）を操作することに応答して、プライマーとする領域を選択する処理を実行するプライマー作成支援部 20 とを有している。

#### 【0019】

プライマー設計支援システム 10 は、コンピュータに、設計支援プログラムをインストールすることにより実現できる。また、本実施の形態においては、ゲノム配列データベース（DB）22 からゲノム配列を読み出すようになっている。ゲノム配列 DB 22 は、上記パーソナルコンピュータのハードディスク上にあっても良いし、或いは、上記パーソナルコンピュータとは離間した配置されたサーバに搭載されていても良い。後者の場合には、パーソナルコンピュータが、LAN やインターネットなどのネットワークを介して、サーバにアクセスして、ゲノム配列 DB 中のデータを参照するように構成すればよい。

#### 【0020】

上記プライマー設計支援システム 10 による処理を説明するのに先立って、本発明の原理について、以下に簡単に説明を加える。

図 2（a）は、ゲノム配列の出現頻度を説明するための図である。ここでは、「ATATGGGATC」という極めて短いゲノム配列を考え、また、部分列として、2-mer 配列を考えた。図 2（a）に示すように、「AT」という 2-mer 配列は、上記ゲノム配列において 3 回出現し、また、「GG」という 2-mer 配列は 2 回出現していることがわかる。他の 2-mer 配列（「TA」、「TG」、「GA」、「TC」）は、それぞれ 1 回ずつ出現し、さらに他の配列（たとえば、「AA」、「AC」など）は一度も出現していないことがわかる。

出現頻度算出部 12 は、このように、2-mer 配列の各々が、ゲノム配列中に何回出現するか（出現頻度）を算出する。

## 【0021】

次に、孤立度について、図2(a)の例と同じゲノム配列および2-mer配列を引用して説明を加える。図3は、上記ゲノム配列に対する各2-mer配列(部分列)の孤立度を説明するための図である。本明細書においては、孤立度を、配列同士の最小ハミング距離として定義した。すなわち、ある $i$ 箇所( $i < n$ )の塩基を別の塩基に入れ替えても、入れ替えられた部分列が、ゲノム配列中に現れないが、 $n$ 箇所の塩基を別の塩基に入れ替えた場合に、入れ替えられた部分列が、ゲノム配列中に出現する場合に、当該部分列の孤立度を「 $n$ 」とする。

## 【0022】

図3に示す例において、部分列「AT」の1箇所の塩基を入れ替えたもの(1ミューテーション: 符号302参照)、つまり、「AA」、「AG」、「AC」、「TT」、「GT」、「CT」は、ゲノム配列(符号300参照)中に表れない。これに対して、2箇所の塩基を入れ替えたもの(2ミューテーション: 符号303参照)、つまり、「TA」、「TG」、「TC」、「GA」、「GG」、「GC」、「CA」、「CG」、「CC」のうち、図3において下線を付したもの(「TA」、「TC」、「GA」、「GG」)は、ゲノム配列(符号300参照)中に表れる。したがって、部分列「AT」の孤立度は「2」となる。

同様に、他の部分列「TA」、「TG」、「GG」、「GA」、「TC」については、その1ミューテーションが、ゲノム配列中に現れるため、それぞれ、孤立度が「1」となる。

## 【0023】

本実施の形態においては、たとえば、18-merの部分列を用いて、出現頻度および孤立度を算出している。図4は、本実施の形態にかかる出現頻度の算出に際して実行される処理の概略を示すフローチャートである。図4に示すように、出現頻度算出部12は、 $N$ -mer配列(たとえば、18-mer配列)を選択して(ステップ401)、ゲノム配列DB22から取得したゲノム配列を走査する(ステップ402)。これにより、ゲノム配列中で、 $N$ -mer配列が出現する位置が特定され、また、これをカウントすることにより出現頻度を得ることができる(ステップ403)。



## 【0024】

出現頻度算出部12は、ステップ403にて得られた出現頻度等を、処理対象となったN-mer配列をキーとして、頻度／孤立度テーブル16中に記憶する（ステップ404）。このような処理を、考えられ得る全てのN-mer配列に関して実行する（ステップ405参照）。このようにしてテーブルを生成することができる。

## 【0025】

実際には、本実施の形態においては、出現頻度が0であるようなN-mer配列を表さないようにすることで、テーブルのサイズを小さくしている。たとえば、図2（a）に示す例において、出現頻度を含むテーブルは、そのままでは、図5（a）に示すようなものとなるが、本実施の形態においては、図5（b）に示すようにサイズを圧縮している。上記テーブルを、場合によって「マップサイズテーブル」とも称する。

## 【0026】

さらに、テーブルの参照箇所を限定する形態を採用することにより、処理の高速化を図っている。たとえば、図2（a）に示す例において、2-mer配列自体をキーとすれば、図5（b）に示すようなテーブル501が得られるが、処理の高速化を図るため、図5（b）に示すように、より小さな長さ（これを、場合によって「ハッシュサイズ」とも称する。）の塩基配列をキーとする補助テーブル502を設け、これにより、テーブル中の配列のおおよその位置を特定できるようにしている。このような補助テーブルを、場合によって、「ハッシュサイズテーブル」とも称する。N=18としたマップサイズテーブルを利用する場合には、たとえば、ハッシュサイズ $\leq 14$ とするのが望ましい。

## 【0027】

次いで、孤立度算出部14は、図4の処理により得られたN-mer配列と頻度とが対応したマップサイズテーブルおよび補助テーブル（ハッシュサイズテーブル）を参照することで、孤立度を算出している。図6は、孤立度の算出処理を示すフローチャートである。

## 【0028】

孤立度算出部 14 は、まず、N-mer 配列を選択する（ステップ 601）とともに、ミューテーションの個数を示す  $i$  を 1 に初期化する（ステップ 602）。次いで、N-mer 配列の  $i$  ミューテーションとなる他の N-mer 配列が選択される（ステップ 603）。孤立度算出部 14 は、上記ハッシュサイズテーブルを参照して（ステップ 604）、他の N-mer 配列が、先頭のハッシュサイズ長の塩基配列が、ゲノム配列中に出現しているか否かを判断する（ステップ 605）。

#### 【0029】

ステップ 605 においてノー(No)と判断された場合に、 $i$  ミューテーションとなる他の N-mer 配列があるときには（ステップ 606 でノー(No)）、当該他の N-mer 配列に関してステップ 603～605 の処理が実行される。或いは、 $i$  ミューテーションとなる全ての他の N-mer 配列について、ゲノム配列における出現の有無が判断された場合には、 $i$  をインクリメントして（ステップ 607）、当該インクリメントされた  $i$  に関する  $i$  ミューテーションについて、同様の処理（ステップ 603～606）を繰り返す。

#### 【0030】

ここに、 $i$  ミューテーションとなる N-mer 配列の特定およびテーブル参照（ステップ 603、604）の手法につき説明を加える。本実施の形態においては、実際に、ステップ 603 においては、ステップ 601 で選択された N-mer 配列の先頭から所定数の塩基が同一であるようなハッシュサイズの配列（ハッシュ配列）を特定し、その塩基のうち、 $i$  個の塩基を変換した配列を生成し、ハッシュサイズテーブル、次いで、マップサイズテーブルを参照して、当該ハッシュ配列を先頭に含む N-mer 配列がどれだけ存在するかを求める。これにより、N-mer 配列のリストが得られる。

#### 【0031】

次いで、孤立度算出部 14 は、得られた N-mer 配列の各々と、処理対象となった N-mer 配列（ステップ 601 で選択されたもの）との間のハミング距離を算出し、ハミング距離の最小値 =  $i$  であるか否かを判断する（ステップ 608）。これは、上記リストされた N-mer 配列には、処理対象となった N-m

e r 配列の  $i$  ミューテーションが全て含まれているため、その最小値が  $i$  であれば、残りの計算をするまでもないからである。

#### 【0032】

ステップ608においてイエス(Yes)と判断された場合には、 $i$  が、処理対象となった  $N$ -mer 配列の孤立度としてテーブルに記憶される。その一方、ステップ608においてノー(No)と判断された場合には、ハミング距離の最小値が  $i$  より大きかった場合、つまり、 $(i+x)$  ミューテーション ( $x \geq 1$ ) が存在することになる。そこで、 $i$  をインクリメントして、ステップ603、604を繰り返して、リストされた  $N$ -mer 配列と処理対象となった  $N$ -mer 配列とのハミング距離の最小値が  $i$  に等しいかどうかを判断する。このようにして、膨大な処理時間を要することなく、各  $N$ -mer 配列の孤立度を算出することができる。

#### 【0033】

可視化処理部18は、上記図4および図5の処理により得られた頻度／孤立度テーブル16を利用して、ゲノム配列中の各塩基を可視化して画像を生成する。この手法につき以下に説明を加える。たとえば、本実施の形態においては、上記  $N$ -mer 配列の出現頻度に基づいて、塩基ごとに出現頻度に関する指標を得るようにしている。図2(b)は、 $N=6$ である場合に、あるゲノム配列(符号211)中の要素「A(矢印212参照)」の指標の算出法を例示する図である。ここでは、要素「A」を含むような  $N$  個の配列を考え、( $N$ 個の配列の出現頻度の総和)／ $N$ を算出することにより第1の指標を求めている。

#### 【0034】

図2(b)の例では、6-mer 配列「ATGCCA」、「TGCCAG」、「GCCAGT」、「CCATGC」、「CAGTCA」および「AGTCAG」が、ゲノム配列中に、それぞれ、8回、2回、3回、1回、3回および4回出現している。したがって、求めるべき指標は、 $(8+2+3+1+3+4)/6$ となる。このような指標の算出は、頻度／孤立度テーブルを参照することにより、高速に実現できる。

また、孤立度についても同様の手法で、塩基ごとの孤立度に関する第2の指標

を求めることが可能である。この場合にも頻度／孤立度テーブルを参照すれば、高速な指標の算出を実現することができる。

#### 【0035】

可視化処理部18は、たとえば、上記第1の指標、第2の指標、或いは、第1の指標および第2の指標を組み合わせた第3の指標に基づいて、ゲノム配列を表示する際の各塩基に割り当てる色彩を決定する。本実施の形態によれば、出現頻度が小さければ小さいほど、つまり、第1の指標の値が小さいほど、当該塩基を含む配列がユニークなプライマーである可能性が高く、その一方、孤立度が大きいほど、つまり、第2の指標の値が大きいほど、当該塩基を含む配列がユニークなプライマーである可能性が高い。これらを利用して、第3の指標＝（第2の指標／第1の指標）として、第3の指標が大きいほど、当該塩基を含む配列がユニークなプライマーである可能性が高くなるように設定しても良い。

#### 【0036】

可視化処理部12は、図7に示すように、第1の指標の値が小さいほど寒色となり、値が大きくなるのに従って暖色となるように設定し、或いは、第2の指標の値が大きいほど寒色となり、その値が小さくなるのにしたがって暖色となるように設定して、当該設定にしたがって、塩基ごとに色を割り当てる。上述した第3の指標については、図7に示すように、その値が大きいほど寒色になることが理解できる。

#### 【0037】

このように、ゲノム配列の塩基ごとに出現頻度および／または孤立度を考慮した色彩が付与されたような画像が表示装置24の画面上に表示されると、オペレータはこれを参照して、よりユニークであると考えられるプライマーの候補を特定することができる。ユーザは、ゲノム配列に付された色彩を参照することで、直感的に、ユニークであると考えられるプライマーの候補を見出すことが可能となる。プライマー作成支援部20は、配列内の相補鎖の形成の有無や、融解温度の選択をなすためのツール、最適なGC含有量、短いリピート配列や回文配列を避けるためのツール（プログラム）を含み、ユーザが選択したプライマー候補について、ユーザの指示にしたがって必要な処理を施すことが出来るようになって

いる。このようにして、ユーザは所定のプライマーをデザインすることが可能となる。

#### 【0038】

本実施の形態によれば、所定長（N）の配列のゲノム配列中の出現頻度、および、当該ゲノム配列に対する上記所定長の配列の孤立度に基づいて、ゲノム配列を構成する各塩基を可視化している。したがって、ユーザはよりユニークな塩基を含むような配列を直感的に視認することが可能となる。また、出現頻度および孤立度の計算のために、頻度／孤立度テーブルを利用することにより、上記可視化に要する処理時間を短縮している。さらに、所定長の配列に関する孤立度を生成する際に、上記Nよりも短い長さの配列に関するハッシュテーブルを利用することで、処理時間の短縮化を図っている。

#### 【0039】

本発明は、以上の実施の形態に限定されることなく、特許請求の範囲に記載された発明の範囲内で、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

たとえば、前記実施の形態においては、出現頻度に関するテーブルにおいて、所定長（N）の配列に関するマップサイズテーブルと、それより短い長さの配列に関するハッシュテーブルの双方を生成している。これを利用して、たとえば、所定長の配列に関する孤立度の算出や、指標の生成に際する出現頻度の特定を実現している。しかしながら、このような構成に限定されるものではなく、マップサイズテーブルのみを備え、いわゆる2分探索で処理を実行しても良い。

#### 【0040】

また、前記実施の形態において、マップサイズを18（N=14）として、かつ、ハッシュサイズ $\leq 14$ としているが、このようなものに限定されず、他のサイズの配列に関してテーブルを生成しても良いことは言うまでもない。

さらに、塩基ごとの指標も上記実施の形態のものに限定されず、また、指標に基づく可視化の手法も、上記実施の形態のものに限定されないことも明らかである。

#### 【0041】

また、前記実施の形態においては、指標に応じて異なる色彩を割り当てているが、これに限定されるものではなく、グレイスケールの異なる明度を割り当てても良いし、さらに、指標に応じて異なる表示態様を割り当てても良い。

#### 【0042】

さらに、前記実施の形態において、プライマー設計支援システムは、出現頻度算出部12および孤立度算出部14を備え、ゲノム配列DB22からの配列に基づいて、頻度や孤立度を示すテーブルを生成している。また、生成されたテーブルを可視化処理部18が利用している。しかしながら、これら全てを備えている必要はなく、たとえば、出現頻度算出部12および孤立度算出部14を備えたシステムによりテーブルが作られ、当該テーブルがCD-ROMやDVD-ROMなどの記録媒体に記録されても良い。この場合には、可視化処理部18を有するシステムが、上記記録媒体を読み取り、塩基に関する指標に沿って異なる色彩を割り当てるなどの処理を実行すればよい。

また、本明細書において、一つの手段の機能が、二つ以上の物理的手段により実現されても、若しくは、二つ以上の手段の機能が、一つの物理的手段により実現されてもよい。

#### 【0043】

##### 【発明の効果】

本発明によれば、ユニークなプライマーの設計を支援することが可能となる。

##### 【図面の簡単な説明】

【図1】 図1は、本発明の実施の形態にかかるプライマー設計支援システムの概略を示すブロックダイヤグラムである。

【図2】 図2(a)は、本実施の形態において、ゲノム配列の出現頻度を説明するための図、図2(b)は、出現頻度に基づく第1の指標を説明するための図である。

【図3】 図3は、本実施の形態における孤立度を説明するための図である。

【図4】 図4は、本実施の形態にかかる出現頻度算出の処理を示すフローチャートである。

【図5】 図5は、本実施の形態にかかる出現頻度に関するテーブルを説明す

るための図である。

【図 6】 図 6 は、本実施の形態にかかる孤立度算出の処理を示すフローチャートである。

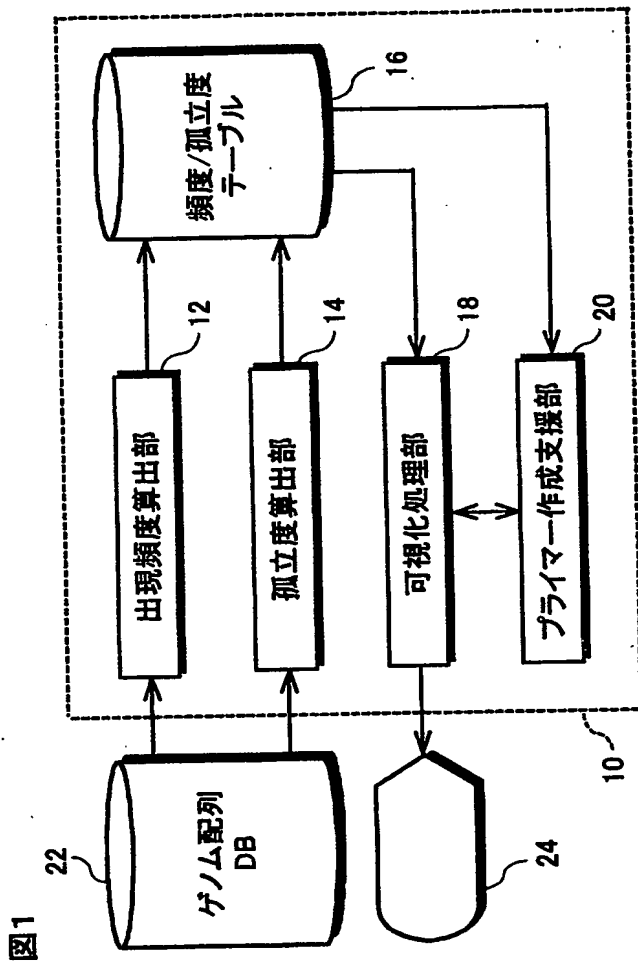
【図 7】 図 7 は、本実施の形態にかかる可視化処理にて割り当てられる色彩を説明するための図である。

【符号の説明】

- 1 0 プライマー設計支援システム
- 1 2 出現頻度算出部
- 1 4 孤立度算出部
- 1 6 頻度／孤立度テーブル
- 1 8 可視化処理部
- 2 0 プライマー作成支援部
- 2 2 ゲノム配列 D B
- 2 4 表示装置

【書類名】 図面

【図1】





【図2】

図2

(a)

ゲノム配列:	ATATGGGATC	出現頻度
N-mer配列:	AT	AT:3
	TA	TA:1
	AT	AT:3
	TG	TG:1
	GG	GG:2
	GG	GG:2
	GA	GA:1
	AT	AT:3
	TC	TC:1

200

(b)

211	...ATGCCAGTCAG...	出現頻度
	ATGCCA	8
	TGCCAG	2
	GCCAGT	3
	CCAGTC	1
	CAGTCA	3
	AGTCAG	4

210

212

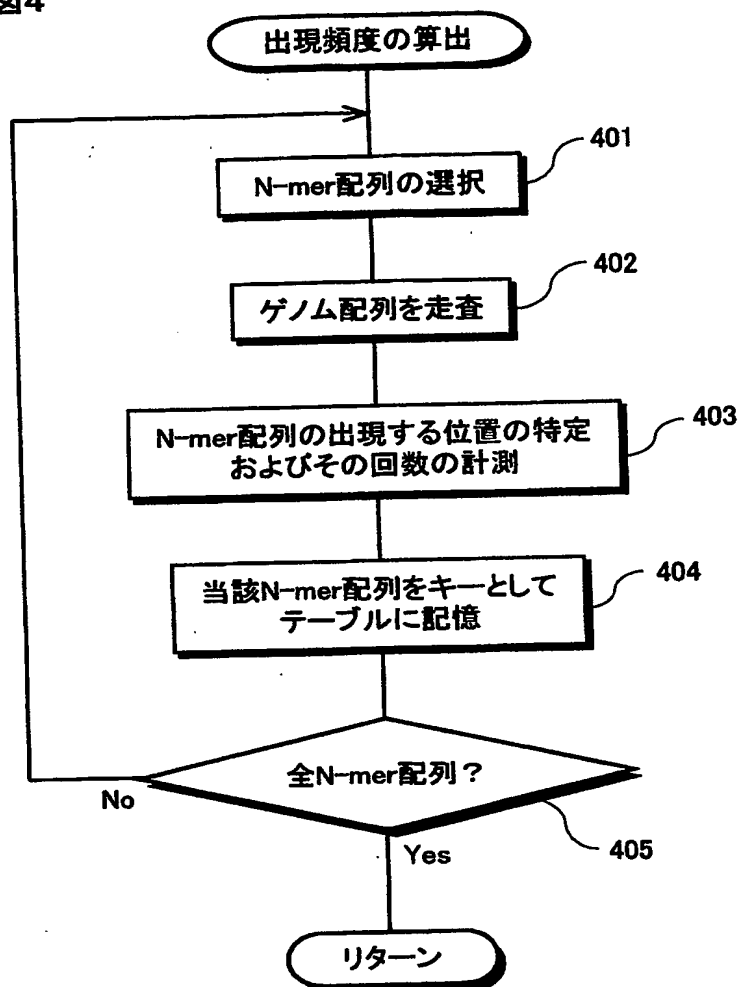
$(8+2+3+1+3+4)/6$ 回出現

【図 3】

図 3	300	301	302	303	304
	ATATGGGATC	出現頻度	1 ミュージョン	2 ミュージョン	孤立度
	AT	AT: 3	AA, AG, AC, TT, GT, CT	TA, TG, TC, GA, GG, GC, CA, CG, CC	2
	TA	TA: 1	AA, GA, CA, TT, TG, TC		1
	AT	AT: 3	AA, AG, AC, TT, GT, CT	TA, TG, TC, GA, GG, GC, CA, CG, CC	2
	TG	TG: 1	TA, TC, TT, AG, CG, GG		1
	GG	GG: 2	GA, GC, GT, AG, CG, TG		1
	GG	GG: 2	GA, GC, GT, AG, CG, TG		1
	GA	GA: 1	GG, GC, GT, AA, CA, TA		1
	AT	AT: 3	AA, AG, AC, TT, GT, CT	TA, TG, TC, GA, GG, GC, CA, CG, CC	2
	TC	TC: 1	TA, TG, TT, AC, GC, CC		1

【図4】

図4



【図5】

図5

(a)

N-mer配列	TT	TC	TG	TA	CT	CC	CG	CA	GT	GC	GG	GA	AT	AC	AG	AA
出現頻度	0	1	1	1	0	0	0	0	0	0	2	1	3	0	0	0

(b)

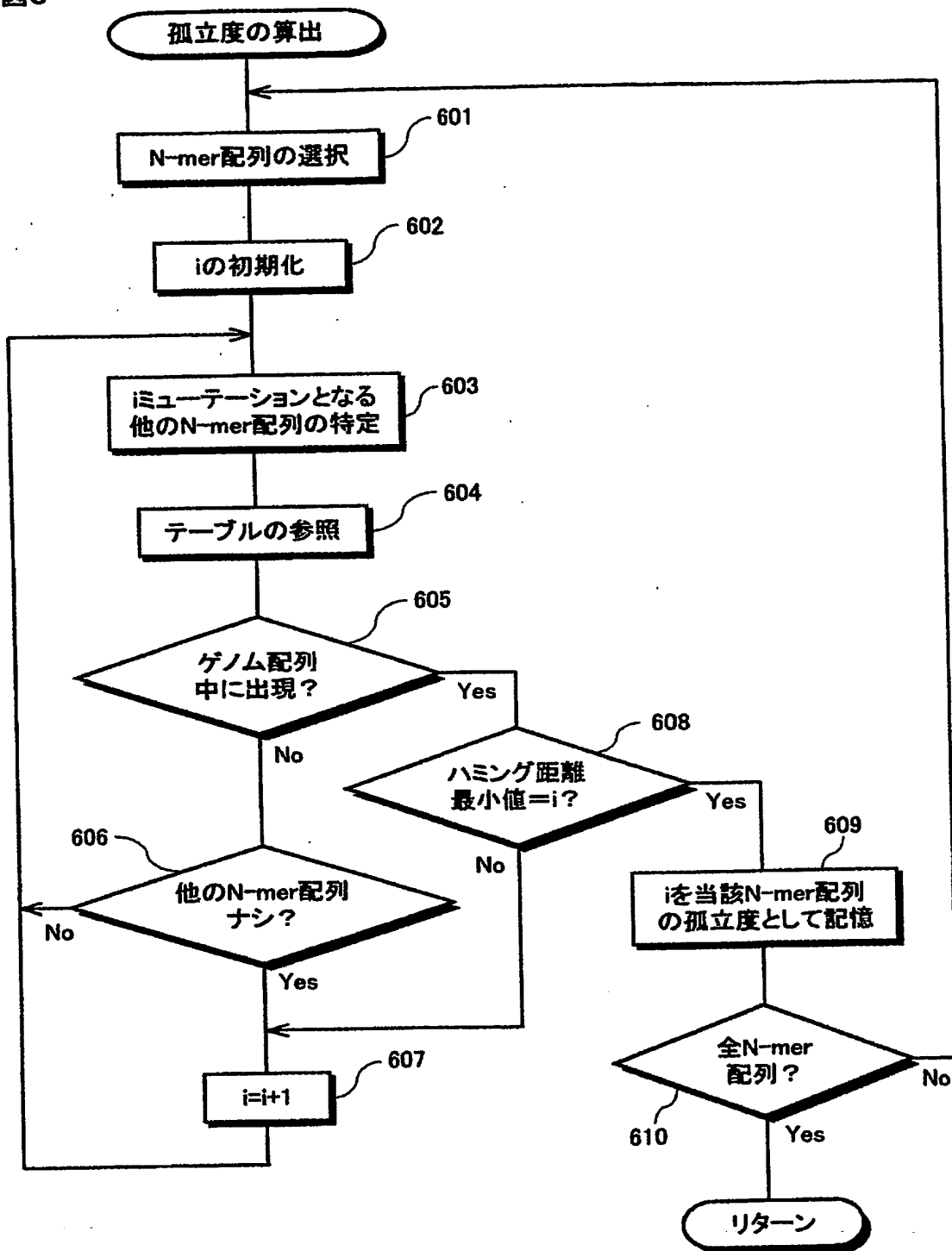
N-mer配列	TC	TG	TA	GG	GA	AT
出現頻度	1	1	1	2	1	3

(c)

ハッシュ配列	T	C	G	A
出現場所	1	0	4	6

【図 6】

図6



【図7】

図7

第1の指標 (出現頻度)	第2の指標 (孤立度)	第3の指標 (第2の指標/第1の指標)	色彩
大 ↑ ↓ 小	小 ↑ ↓ 大	小 ↑ ↓ 大	暖色 ↑ ↓ 寒色

【書類名】 要約書

【要約】

【課題】 ユニークなプライマーの設計を支援する。

【解決手段】 ゲノム配列中の配列の出現頻度を示す指標を算出するために、  
所定長の部分列の各々に関して、前記ゲノム配列中の出現頻度を算出し、所定長の部分列に関する出現頻度を、頻度／孤立度テーブル16に記憶する。また、所定長の部分列の各々に関して、 $j$  個の塩基を変換した  $j$  ミューテーション ( $j \leq i-1$ ) が、ゲノム配列中には表れないが、 $i$  個の塩基を変換した  $i$  ミューテーションが、ゲノム配列中に表れるような孤立度  $i$  を算出し、所定長の部分列に関して、孤立度を、頻度／孤立度テーブル16に記憶する。可視化処理部18は、塩基の出現頻度および／または孤立度に基づき、塩基の各々に所定の色彩を付与してゲノム配列を表したような画像を生成する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願2001-396194
受付番号	50101911040
書類名	特許願
担当官	第七担当上席 0096
作成日	平成14年 1月24日

<認定情報・付加情報>

【提出日】	平成13年12月27日
-------	-------------



【書類名】 出願人名義変更届  
【あて先】 特許庁長官殿  
【事件の表示】  
    【出願番号】 特願2001-396194  
【承継人】  
    【識別番号】 899000024  
    【氏名又は名称】 株式会社先端科学技術インキュベーションセンター  
    【代表者】 山本 貴史  
【承継人代理人】  
    【識別番号】 100103632  
    【弁理士】  
    【氏名又は名称】 窪田 英一郎  
【承継人代理人】  
    【識別番号】 100099715  
    【弁理士】  
    【氏名又は名称】 吉田 聡  
【手数料の表示】  
    【予納台帳番号】 058377  
    【納付金額】 4,200円  
【プルーフの要否】 要

認定・付加情報

特許出願の番号	特願2001-396194
受付番号	50201114008
書類名	出願人名義変更届
担当官	土井 恵子 4264
作成日	平成14年 9月 9日

<認定情報・付加情報>

【提出日】 平成14年 7月29日

【承継人】

【識別番号】 899000024

【住所又は居所】 東京都千代田区丸の内一丁目5番1号 新丸の内  
ビルディング6階

【氏名又は名称】 株式会社 先端科学技術インキュベーションセン  
ター

【承継人代理人】 申請人

【識別番号】 100103632

【住所又は居所】 東京都港区赤坂2丁目21番8号 赤坂山田ビル  
1階 窪田・吉田法律特許事務所

【氏名又は名称】 窪田 英一郎

【承継人代理人】

【識別番号】 100099715

【住所又は居所】 東京都港区赤坂2丁目21番8号 赤坂山田ビル  
1階 窪田・吉田法律特許事務所

【氏名又は名称】 吉田 聡

出 願 人 履 歴 情 報

識別番号

[500524969]

1. 変更年月日 2000年12月13日

[変更理由] 住所変更

住 所 神奈川県横浜市都筑区南山田1-3-4-1202

氏 名 森下 真一

出 願 人 履 歴 情 報

識別番号

[502001433]

1. 変更年月日 2001年12月27日

[変更理由] 新規登録

住 所 東京都文京区本郷5-14-8-302

氏 名 山田 智之

出 願 人 履 歴 情 報

識別番号 [899000024]

1. 変更年月日 1999年 9月16日

[変更理由] 新規登録

住 所 東京都千代田区丸の内一丁目5番1号 新丸の内ビルディング6階

氏 名 株式会社 先端科学技術インキュベーションセンター